Prediction in second language processing and learning: Advances and directions

Edith Kaan¹ and Theres Grüter²

¹University of Florida, ²University of Hawai‘i at Mānoa

Version date: 03/09/2021

This article is to be published as:


Contact: kaan@ufl.edu; theres@ufl.edu
Abstract

There is ample evidence that language users, including second language (L2) speakers, can predict upcoming information during listening and reading. Yet it is still unclear when, how, and why language users engage in prediction, and what the relation is between prediction and learning. This volume presents a collection of current research, insights and directions regarding the role of prediction in L2 processing and learning.

In this introductory chapter, we provide an overview of the current state of the field and highlight that prediction may not always be the most efficient processing mechanism, depending on a language user’s linguistic experience, task demands, goals and resources. We propose that a focus on the utility of prediction may help us better understand differences in predictive processing within and between individuals and groups.

Introduction

There is ample evidence that language users can predict upcoming information during listening and reading. For instance, many studies using visual world paradigms have shown that listeners can move their eyes to a particular object in the display even before the object is mentioned, suggesting they can develop clear expectations of how the sentence
continues (Altmann & Kamide, 1999). According to some approaches, predictive processing is ubiquitous and essential for learning (Chang et al., 2006; Dell & Chang, 2014). Following the general trend in sentence processing research, predictive processing has become a popular topic in second-language (L2) processing and acquisition research as well. While several initial L2 studies failed to find evidence for predictive processing in L2 learners (Grüter et al., 2012; Lew-Williams & Fernald, 2010; Martin et al., 2013), there now seems to be consensus that L2 users can predict. However, we are still in need of a better understanding of when, how, and why L2 (and first language!) comprehenders engage in prediction, and what the relation is between prediction and learning.

The goal of this volume is to provide an overview of current research directions and recent insights related to prediction in L2 processing and learning. We will start this introductory chapter by addressing different views of prediction and its hypothesized role in language processing and learning. We then give a brief overview of past and current developments in predictive processing research in first and second-language. We advocate an approach taking into account the utility of prediction (Kuperberg & Jaeger, 2016): Prediction may not always be the most efficient way to process language, depending on various factors. We believe that this approach may push research forward into explaining differences in predictive processing and learning within and across
individuals and groups. We then summarize the chapters in this volume, and conclude by sketching directions for further research.

What is prediction and what is it good for in L2 learning and processing?

What is prediction?

The term “prediction” has been interpreted in different ways in different research traditions in psycholinguistics and cognitive science. This makes the discussion about the function of predictive processing, and the evidence for prediction not always clear (DeLong et al., 2014; Kuperberg & Jaeger, 2016). A very wide conceptualization of “prediction” is that the “context influences the state of the language processing system before the bottom-up input is observed” (Kuperberg & Jaeger, 2016, p. 33). The “state of the language processing system” can be influenced by many factors, linguistic as well as non-linguistic, and changes with every new input. Aspects of the new input (semantics, form, and others) can therefore be more or less expected depending on the prior state of the processing system. As a result, the actual new input can be easier or harder to process.
In line with this wide view of prediction, Ferreira and Chantavarin (2018) propose the term “preparedness”: a state of readiness for new information constrained by the context. According to Ferreira and Chantavarin, the function of predictive cues is to signal upcoming new information that is to some extent restricted by the preceding context (Myslíň & Levy, 2016). This in turn will make it easier for the reader/listener to process (integrate) this new information. In this view, prediction and integration are two sides of the same coin. In addition, this stance implies that prediction of a very specific lexical form is rare in daily communication. The goal of communication is to provide new information. Words that are completely predictable do not convey new information, and hence will not occur often.

A more restrictive view of “prediction”, commonly adopted in sentence processing, is that of pre-activation of various aspects of upcoming information. Given this concept of prediction, relevant research questions are: To what extent can specific words, including low-level features such as phonological or orthographic form, be pre-activated (e.g. Nieuwland, 2019)? Researchers adhering to this view of prediction hold that only effects observed before the onset of the critical word, or in very early stages of the processing of the target word, provide empirical evidence for prediction (Pickering & Gambi, 2018). Measures of processing difficulty at the target word or off-line measures are mostly
inconclusive in this view, since such data can also be explained by re-
active integration processes (Brouwer et al., 2017). Many researchers
adhering to this view also distinguish pre-activation by (automatic, lexical)
associations from more precise predictions constrained by the combination
of various sources of top-down information (Ito & Pickering, this volume;
Pickering & Gambi, 2018; Pickering & Garrod, 2013).

In sum, there is agreement that predictive processing involves
some form of preparedness, expectancy or pre-activation within the
language processing system. Yet views differ in whether this means that
specific forms are pre-activated, and in what is taken to constitute
evidence for prediction. What is uncontroversial is that language users can
and do engage in proactive, forward-looking processing during language
comprehension.

*What is prediction good for?*

Why do readers and listeners predict? Many assume that prediction makes
communication more efficient because it makes upcoming information
easier to process. The functions of these forward-looking aspects of
communicative behavior have been of long-standing interest in the study
of conversational turn-taking, following Sacks et al.'s (1974) observation
that interlocutors predict, or project, the end of the current speaker's turn
Another function ascribed to predictive processing is its role in adaptation and learning. According to error-based implicit learning models (Chang et al., 2006), language users make predictions regarding upcoming information based on their prior experience. When these predictions are not borne out, the difference between the predicted and actual input (“prediction error”) is then used to update the language processing system, such that future prediction errors can be minimized. Belief-updating models (Jaeger & Snider, 2013) are equivalent to error-based learning models in that the language processing system is changed (updated) with each incoming word in such a way that future changes are as small as possible. Error-based implicit learning and belief-updating can account for rapid adaptation by the listener and reader to, for instance, a particular way in which a particular person speaks, or to particular sentence structures used (Fine & Jaeger, 2013; Kleinschmidt et al., 2015). Error-based learning mechanisms have also been proposed to be a part of language learning (Chang et al., 2006; Phillips & Ehrenhofer, 2015). A learner makes predictions. If these predictions are not borne out, the difference between the predicted and actual input serves as implicit corrective feedback, pushing the learner more towards the target language. This proposal critically rest on the assumption that language learners can and
do make predictions, and that they are able to use the implicit feedback provided by the actual input to update their processing system in case of a prediction error. These assumptions are not uncontested in the context of (L2) learning (Huettig & Mani, 2016; Kaan, 2015; Kaan et al., 2019; see also Hopp, this volume; Gambi, this volume).

Others have questioned the importance of prediction, pointing out that prediction is not always observed and is not essential for language processing and learning (Huettig & Mani, 2016). After all, most studies reporting evidence of prediction use highly constraining contexts, which may not be representative of natural language use. Moreover, there are phenomena that cannot be learned through prediction and mechanisms other than prediction through which learning occurs (Huettig & Mani, 2016; Kaan, 2015; see also Hopp, this volume). This suggests that prediction is not essential to processing and learning.

In this chapter we will suggest a way to reconcile the debate concerning the scope of prediction and its importance. We will start with a brief history of research on predictive processing.

A brief history of prediction in language processing research
Predictive processing has not always been a popular topic of research. In the 1980s and 1990s, modular conceptions of the mind typically stressed bottom-up flow of information (that is, from sensory perception to “higher” levels of cognition), with top-down information only serving as a way to select among multiple options provided by the bottom-up input. As a consequence, language processing research gave priority to bottom-up processing of perceptual and word-level information, with the role and timing of higher-level discourse and non-linguistic information being heavily debated. Predictive sentence processing was considered not very feasible: A sentence could end in an infinite number of possible ways, so predictions would often be wrong and therefore costly (see Ferreira & Chantavarin, 2018, for a concise historical review).

Around the first decade of the 21st century, cognitive science experienced a shift in thinking about information flow in the mind and brain. Friston (2010) and Clark (2013) argued for a view in which top-down information was of primary importance, with bottom-up information serving only a corrective role. In this view, the mind and brain are seen as “prediction machines” (Clark, 2013; Van Berkum, 2010). At any level of processing, predictions are made as to what will come next. These predictions are compared with what actually occurs. The difference between prediction and input is new information, which is used to fine-
tune knowledge and processes at multiple levels. In this way, humans and other organisms can optimally adjust to the dynamic world around them.

*Research on prediction in L1 sentence processing*

In the realm of sentence processing, new experimental techniques such as visual world eyetracking (Tanenhaus et al., 1995), and event-related brain potentials (Kaan, 2007) provided evidence for predictive processing. Altmann and Kamide (1999) conducted an eye-tracking study showing that listeners can anticipate upcoming information: Listeners can move their eyes to a specific object in a display even before this object is verbally mentioned, that is, listeners look more at the edible object in the display when they hear *The boy will eat the…* compared to *The boy will move the…*. Also results from studies using event-related brain potentials (ERPs) suggested that language users can predict upcoming words. A seminal study by DeLong et al. (2005) used highly constraining contexts such as *It was a windy day so the boy decided to fly a/an…*. The ERPs showed incongruency effects at *alan* when the form was not compatible with the most likely completion (e.g., when the form was *an* even though the most likely completion was *kite*). The size of the incongruency effect at the determiner correlated with how likely a completion the noun was. Note that these effects were seen before the noun was presented. This
supported the view that prediction is gradual, and that language users predict upcoming words even to the level of the phonological form (that is, whether a word starts with a vowel or consonant, and hence requires *an* or *a*). Even though the replicability of these results is still being debated (Nicenboim et al., 2020; Nieuwland et al., 2018), studies using gender-marked languages such as Spanish and Dutch rather consistently found effects of incongruency when the gender marking on the determiner or adjective did not match that of an expected but yet-to-be-presented noun, suggesting that specific word features can be predicted, including gender information (Foucart et al., 2014; Foucart et al., 2016; Van Berkum et al., 2005; Wicha et al., 2004). Many studies conducted in recent years have shown that (adult L1) language users can predict, and can predict many types of information (lexical, thematic, phonological, syntactic, semantic, referential) on the basis on many types of cues (gender, case, discourse, prosody, verb form, verb meaning, among others); for a recent overview, see Pickering and Gambi (2018). Nevertheless, it remains under debate just how ubiquitous prediction is in adult L1 processing under various task demands and outside the laboratory (e.g. Huettig & Guerra, 2019; Pickering & Gambi, 2018).

*Research on prediction in L2 sentence processing*
Research on L2 sentence processing was initially mainly concerned with differences between L2 learners and native speakers and whether these differences could be accounted for by differences in representation (knowledge, e.g. Hawkins & Chan, 1997), processing and resources (Hopp, 2010), or both (e.g. Clahsen & Felser, 2006). Differences at the level of "processing" were conceived in terms of L2 users' potential limitations in incremental information integration during listening and reading (e.g. Sorace, 2011). However, following the growing interest in predictive processing in the early 2000s, studies soon appeared on predictive processing in L2 learners. Early studies (Grüter et al., 2012; Lew-Williams & Fernald, 2010; Martin et al., 2013) failed to find evidence for prediction in late second language learners. This led to views expressed in Kaan et al. (2010) and Grüter et al. (2014), hypothesizing that L2 learners have no or a Reduced Ability to Generate Expectations (RAGE). However, it quickly became clear that L2 speakers can make predictions during sentence comprehension. Even learners with very limited experience were shown to be able to use semantic information, such as verb restrictions, to anticipate a noun (Koehne & Crocker, 2014), and some L2 speakers were like native speakers in using gender information to predict upcoming nouns (Dussias et al., 2013; Hopp, 2013). It is therefore clear that the ability to predict in an L2 is not an all-or-nothing issue. In other words, the question is not whether L2 users can or
cannot predict (they can), but what factors and circumstances affect the extent to which L2 users engage in predictive processing.

In a programmatic overview, Kaan (2014) outlined sources of variation in prediction. Many of these sources of variation are shared between native and L2 speakers (frequency, lexical quality, resources, task-induced processes and strategies, among others). Kaan (2014) therefore argued that L1 and L2 processing are not different in terms of predictive machinery, and that research should be aimed at identifying and quantifying sources of variation that drive predictive processing. Many studies on L2 prediction, including several in this volume, follow this line of inquiry.

A second, more recent line of research concerns the role of prediction in L2 learning. According to error-based learning models, prediction is a critical component of learning: If what you see or hear is different from what you expect, you will use this error to adjust your knowledge and future expectations. However, if L2 speakers engage in prediction less or differently from native speakers, how exactly can learning from prediction errors contribute to L2 learning? This complex question has only recently begun to be addressed (Hopp, this volume; Jackson & Hopp, 2020; Kaan & Chun, 2018a).

In what follows, we will highlight some recent insights that we see as particularly relevant to predictive processing in L1 and L2 and its
relation to learning, and advocate an approach that focuses on the utility of prediction (Kuperberg & Jaeger, 2016).

**Prediction in L1 and L2 processing: the role of utility**

*Cue reliability and utility in L1 predictive processing*

One major insight from L1 prediction research is that predictive processing is highly variable even *within* individuals, and that language users adapt their predictive behavior to particular situations, both as to *what* they are predicting and *whether* they are predicting. This ties in with observations and ideas concerning within-subject variability in other aspects of language processing (e.g. Ferreira & Patson, 2007; Hsu & Novick, 2016), and other cognitive functions (e.g. Botvinick et al., 2001; Kan et al., 2013). Prediction can be affected by many interrelated factors. We will discuss cue reliability, and goals and other task demands.

Predictions are made on the basis of cues. For instance, gender marking can be a cue for an upcoming type of noun; verb semantics can be cue for an upcoming type of object. Predictions are based on a reliable relation between the cue and what it signals in prior experience. Studies have shown that when cues are no longer reliable in the recent
experimental context, participants stop using them predictively (Brothers et al., 2019; Heyselaar et al., 2020; Hopp, 2016; but see Zhang et al., 2019). For instance, if native German speaking participants are exposed to input that contains occasional errors in gender marking, they no longer use gender-marking as a predictive cue in the context of the study (Hopp, 2016). The reliability of a cue is also affected by the composition of the stimulus materials. For instance, when the proportion of related word pairs is high, participants show a larger and earlier relatedness effect in the ERPs than when the proportion is small, suggesting that participants engage more in prediction in the former case (Lau et al., 2013; see also Coulson et al., 1998; Hahne & Friederici, 2002 for comparable observations related to syntactic violations).

A cue can be considered more or less reliable in combination with other cues, or can change in what it is taken to be predictive of. Prosodic cues can enhance the predictive use of case markers in native German speakers (Henry et al., 2017). Listeners can change their predictions depending on who they are listening to and their prior experience with that talker (Kamide, 2012; Ryskin et al., 2020). Similarly, a talker’s accent can affect predictive behavior. Listeners predict more slowly and to a lesser extent when listening to foreign accented than to native speech, especially when they have less experience listening to the foreign accent (Porretta et al., 2020; Romero-Rivas et al., 2016). This can be explained by the
predictive cue being less reliable, either because it is harder or takes longer to perceive words in a foreign accent, or by the uncertainty of the predictions associated with the cue or speaker (see also Bosker et al., 2019).

The language user’s goals and awareness can also affect prediction. Participants are more likely to predict when instructions explicitly direct them to do so (Brothers et al., 2017), and are more likely to make predictive eye-movements when they are aware of a predictive relation between the cue and the target (Andringa, 2020; Curcic et al., 2019). Another, related set of factors are task demands. In a visual world paradigm, participants are more likely to make predictive eye-movements if they have more time to look at the visual display in advance of the spoken information, and if the speech they listen to is presented at a slower rate (Ferreira et al., 2013; Huettig & Guerra, 2019). Ito et al. (2018) found that predictive eye-movements were delayed when participants had to memorize a set of unrelated words while listening to sentences. This suggests that predictive behavior can be modulated by concurrent memory load (see also Chun et al., this volume). Finally, predictive processing is affected by individual-difference factors such as reading ability, working memory, processing speed, age, and L1 exposure (Huettig & Brouwer, 2015; Huettig & Janse, 2015; Huettig & Pickering, 2019; Ito & Sakai, 2021; Mishra et al., 2012; Wlotko et al., 2012).
The above observations suggest that there is a lot of variability in prediction and in the way prediction occurs, even within individuals. Kuperberg and Jaeger (2016) therefore stress the role of utility: Language users weight the cost of prediction against the benefits to optimally reach their communicative goals. In some cases, using certain cues to predict will lead to too many errors or will exceed resources, and hence will be costly; in this case, depending on their goals, some listeners or readers may stop using this cue predictively, rely on a different cue, or may not predict at all in order to achieve optimal efficiency. Individuals therefore dynamically adjust their predictive behavior (what they predict and whether they predict) depending on the reliability of the predictive cues, their goals, task demands, resources, and other factors.

*Cue reliability and utility in L2 predictive processing*

As pointed out by Kaan (2014), the same sources that underlie variation in prediction in L1 may account for much of the variability we see in predictive processing in L2 learners and other bilinguals. Everything discussed in the previous section therefore applies to L2 learners as well. There are, however, some potential sources of variation that are unique to, or at least more pronounced, in L2 learners and bilinguals in general: cross-linguistic influence (competition or transfer between languages),
proficiency, and exposure, including the way the L2 is learned. Under the view we advocate, these sources factor into cue reliability, cue weighting, and, consequently, the utility of prediction. The calculation of cost-benefit trade-offs may not always yield the same result in an L2 as in an L1. As a result, there may be circumstances where prediction is the most efficient processing option for L1 speakers but not for L2 speakers, or L2 speakers may make predictions that are different from those made by L1 speakers.

L2 users may place different weights on different cues, in line with the differential reliability of these cues within their language systems. According to the Competition Model (Bates & MacWhinney, 1981; MacWhinney & Bates, 1989) languages weigh cues differently. For instance, in English, word order is a reliable cue to determine the subject and object of a clause, whereas in Spanish subject-verb agreement is more important. What is a reliable cue for prediction in the L1 may more easily be regarded as a reliable cue in the L2 if these cues are shared between the languages. Evidence that reliable cues can be transferred from the L1 to the L2 comes from van Bergen and Flecken (2017). In this study, L2 speakers of Dutch anticipated the orientation of an object based on the information encoded in the verb, but only if the learners’ L1 also encoded object orientation in placement verbs. This suggests that semantic information encoded in the verb transfers from one language to the other and can be used predictively.
Morphosyntactic cues that overlap between L1 and L2 may also be used for prediction more, or more easily, in the L2. Dussias et al. (2013) report that intermediate-proficiency English L2 learners of Spanish did not use gender information predictively in Spanish, whereas proficiency-matched Italian learners of Spanish did, at least for feminine gender. This suggests that having a gender-marked L1 can help use gender predictively in a gender-marked L2 (Spanish). This especially holds when specific cues or constructions overlap (Hopp, this volume; Hopp & Lemmerth, 2018; Morales et al., 2015), suggesting that transfer of predictive use of a cue (e.g., gender) is not an all-not-nothing issue, but a function of what overlaps in the languages (see Foucart, this volume).

When there is no straightforward overlap in predictive use of cues between the languages, some cues may not be reliable for an L2 speaker because their representations are not sufficiently specified or entrenched. An example is the use of gender as a cue for a noun in beginning English L2 speakers of Spanish or German. If unreliable cues were used predictively, too many prediction errors would result. Hopp (2016) proposes that this situation is similar to that of L1 speakers who stop using gender predictively when they encounter a speaker who uses gender unsystematically. For an L2 learner, not predicting may be more efficient than generating predictions on the basis of unreliable cues.
Alternatively, L2 learners may interpret cues differently. For example, they may interpret a contrastive pitch accent as a lexical rather than an information-structural cue (Lee et al. 2020), or weigh cues differently, resulting in predictions that are different from those made by native speakers of the language. Grüter et al. (2020) investigated the predictive use of classifiers in Mandarin Chinese by L1 and L2 speakers. While L1 speakers relied on classifiers as a grammatical form class cue, L2 speakers also predicted upcoming nouns on the basis of classifiers but prioritized semantic cues, which are generally (but not perfectly) reliable. Another example is Hopp (2015), who tested English-speaking L2 learners of German with low-intermediate to advanced proficiency. Participants listened to sentences that started with an accusative or nominative noun phrase (e.g., *Den Wolf tötet gleich der Jäger.* ‘the-ACC wolf kills soon the-NOM hunter’, meaning ‘The hunter will soon kill the wolf’, and *Der Wolf tötet gleich den Hirsch.* ‘the-NOM wolf kills soon the-ACC deer’, meaning ‘The wolf will soon kill the deer’). Native German speakers used the case information, reflecting the grammatical function and thematic status of a noun, at the start of the sentence to predict upcoming thematic roles: They predictively fixated on the most likely theme (deer in the example) when the first NP had nominative case, and on the most likely agent (hunter in the example) when the first NP was accusative. In contrast, the L2 speakers consistently made predictive eye-movements to
the likely theme (deer), regardless of the case of the first noun phrase. These results can be accounted for in a similar way: An agent-before-theme strategy is generally very reliable and yields the appropriate interpretation in most cases. Note that when tested off-line, L2 speakers typically perform correctly, so results cannot be explained by a lack of relevant linguistic knowledge. This suggests that in case of uncertainty (due to the processing speed needed when listening, uncertainty regarding what they hear, or due to less specified lexical representations), L2 speakers may rely more on cues they deem reliable from prior experience, that is, will mainly rely on frequent and prototypical associations, especially if these associations can be mapped from L1 to the L2.

Differential reliance on particular cues among L1 and L2 speakers may result from the different manner in which a language is learned in early childhood as opposed to later in life (Arnon & Ramscar, 2012; Grüter et al., 2012; Siegelman & Arnon, 2015). Take the case of grammatical gender. In the L2 classroom, nouns are typically first learned as form-meaning associations; rules pertaining to gender agreement, such as the selection of the correct determiner, are introduced later. By that time, the noun itself constitutes a fully reliable cue to meaning, thus blocking attention to co-occurrence relations with gender-marked determiners. On the other hand, a child acquiring language will hear nouns together with determiners and adjectives, and will form strong associations
between the noun and gender information, hence making gender marking a reliable cue for the prediction of upcoming nouns. This account may in part also explain the difference in prediction on the basis of, e.g., gender between child L1 and L2 learners, and adult L2 learners (see also Karaca et al., this volume).

A note about proficiency
It is perhaps intuitive to assume that L2 speakers' engagement in predictive processing will become more similar to that of L1 speakers as global proficiency increases. Indeed, several studies have reported greater and/or more native-like effects of prediction in groups with higher global proficiency in the L2 (Chambers & Cooke, 2009; Dussias et al., 2013; Henry et al., 2020; Hopp, 2013; Hopp & Lemmerth, 2018; see also Karaca et al. this volume). Notably, however, a substantial number of recent studies that explicitly investigated the correlation between proficiency and prediction found no such effects (Dijkgraaf et al., 2017; Hopp, 2015; Ito et al., 2018; Kim & Grüter, 2020; Mitsugi, 2020; Perdomo & Kaan, 2019), or found that even very highly proficient L2 speakers differ from native speakers in their predictive use of certain features (Dijkgraaf et al., 2019; Kaan et al., 2016). Hopp (2015) report that proficiency did not affect the agent-verb-theme strategy (but cf. Henry et al., 2020). More proficient L2 speakers were however able to revise the initial interpretation more
quickly than less proficient L2 speakers, indicating that the measure of proficiency employed in that study was able to explain relevant variation in processing behavior, but only with regard to information integration/revision, not prediction (see also Mitsugi, 2020). Although the absence of proficiency (or any other effect) needs to be interpreted with caution, observations such as the above suggest that prediction does not automatically come with global proficiency and, vice versa, that high overall proficiency does not imply L1-like prediction. In sum, we propose that, as in L1 processing, prediction in L2 processing is modulated by utility: Generating specific predictions only makes sense in terms of processing efficiency if predictions are based on reliable cues and if predictions are correct most of the time. What is reliable is a function of experience and cross-linguistic influence, and the way the L2 is learned, but may not always be fully explained by global proficiency measures. Since L2 experience differs, it is a natural consequence that L2 speakers may differ from native speakers and from each other in what, when, and the extent to which they predict when processing the L2. Rather than constituting a deficiency, such differences may reflect adaptation to optimize processing efficiency given a language user's available knowledge, experience and resources.

**Prediction and learning**
A second line of research featured in this volume concerns the role of prediction in (L2) learning. As we outlined above, error-based learning models assume that prediction plays a critical role in learning. The difference between what is predicted and the actual input is what gives rise to learning. That is, the processing system is changed as a result of prediction errors to minimize future prediction errors, and thus to more optimally process future input. Error-based learning mechanisms of the kind described have been successfully implemented in computational models of learning (Chang et al., 2006; Rumelhart et al., 1986). However, learning in humans is complicated, and listeners and readers do not exclusively or always use processing errors to adapt to the properties of the language they are exposed to.

One way in which error-based learning has been tested is by exposing participants to sentences with an unusual or infrequent syntactic structure. If readers or listeners adapt to these structures they will show a change in the way they process these structures as a function of exposure. Fine et al. (2013) report evidence indicating that L1 speakers adapt (see also, among others, Jaeger & Snider, 2013; Thothathiri & Snedeker, 2008; Wells et al., 2009). In Fine et al.’s study, participants read sentences
with reduced relative clause ambiguities, such as *The experienced soldiers warned about the dangers conducted the midnight raid.* Fine et al. found the classic garden path effect, that is, reading times were longer at the disambiguating verb (*conducted*) for the temporarily ambiguous condition (example above) versus an unambiguous control sentence (*The experienced soldiers who were told about the dangers conducted the midnight raid*). Assuming a wide conceptualization of prediction, garden path effects can be interpreted as indexing prediction errors (Levy, 2008). Fine et al. found that the size of the garden path effect decreased over the course of the study, suggesting that readers adapted to the reduced relative clause structures. In addition, when two groups were compared, one with and one without prior exposure to reduced relatives, the group with prior exposure to reduced relatives showed smaller garden path effects. Finally, the group who read more reduced relatives eventually showed processing difficulty for simple main clause sentences, suggesting that these readers had reversed their processing preferences as a result of exposure. These findings support the view that participants adapt their processing on the basis of recent exposure. However, other adaptation studies yielded mixed results, with some studies or constructions showing evidence of adaptation but not others (see Kaan & Chun, 2018b for a critical review). For instance Kaan et al. (2019) found that native English speakers adapted to only one of two constructions tested. Hence even though (L1) readers
experience prediction errors as reflected by the garden path effects, they do not always use these errors to adapt or adapt successfully to the structures they are exposed to in the timeframe of the study. The relation between prediction error and changes in the processing system is therefore not straightforward.

*Insights from L2 speakers*

Turning to L2 learning, learning approaches based on prediction error provide a seemingly elegant proposal for how L2 learning could occur through prediction. First, an L2 learner may base their prediction on what is reliable and frequent in their L1. If this prediction is not borne out in the L2, the L2 processing system will then adjust itself to minimize future prediction errors. This adjustment is larger the less expected the new input is. As a result, the learner’s L2 processing system moves away from the L1 and starts to approximate the L2. Evidence supporting this comes from a priming study by Montero-Melis and Jaeger (2020), in which Swedish L2-learners of Spanish listened to directional descriptions and produced descriptions themselves. In Swedish, verbs mainly express the manner of the motion (as in e.g., *to push*), whereas Spanish verbs tend to encode the path of motion. Montero-Melis and Jaeger (2020) found that the less proficient L2 speakers adapted more quickly to the L2 structure that was
infrequent in their L1 but frequent in the L2, whereas the more proficient
speakers adapted more to the L2 structure that was infrequent in the L2.
This can be accounted for in an error-based learning system in which
beginner learners based themselves on the L1 regularities and adapted
their system towards the L2, whereas the more advanced learners had
already moved away from their L1 and fine-tuned the system to match the
distributions of the structures in the L2.

However, the question is to what extent insights from such
laboratory-based learning studies can inform our understanding of
“learning” as typically understood in the context of L2 learning, namely
the acquisition of long-term, generalizable knowledge (Kaan & Chun,
2018b). Jackson and colleagues (Jackson & Hopp, 2020; Jackson & Ruf,
2016) conducted a priming study and found that L2 learners adapted to
structures that were infrequent in the L1 but frequent in the L2 (see also
Kaan & Chun, 2018a). However, the effects disappeared in a post-test,
suggesting that short-term adaptation did not lead to longer-term retention.

Another concern is that L2 speakers, like L1 speakers, do not
always adapt with increased exposure, even though they may experience
prediction errors. In a self-paced reading study, Kaan et al. (2019) found
that both L2 and L1 speakers experienced garden-path effects when a
sentence continued with an infrequent, unexpected structure. The L1
speakers showed a decrease of the garden path effect to one of the
structures over the course of the study, suggesting they had changed their processing as a result of the recent exposure. However, the L2 speakers did not show signs of adaptation to this infrequent structure, suggesting that they did not use the prediction error to adapt, at least, not in the same timeframe (but cf. Hopp, 2020), even though this structure would have been even less frequent to the L2 speakers, and hence would be predicted to lead to more adaptation than in the L1 speakers. These results, on the face of it, are not what one would expect in an error-based learning approach.

_Again: utility_

Why is it that L1 and L2 speakers sometimes do predict but do not adapt or learn? Hopp (this volume) discusses multiple reasons. Here we will focus on the utility of prediction. When a cue is very reliable for a speaker, there may be little or no gain to the speaker in making any changes to their existing system as a result of prediction error. In most cases, predictions will be correct and will contribute to communicative efficiency. Prediction errors induced by occasional exceptions will therefore have little or no effect on the reliability of a cue. In Bayesian terms: If prior information is very reliable, you will not change your beliefs much in the face of rare counterevidence (Kuperberg & Jaeger, 2016).
Aside from cue reliability and communicative efficiency, another factor that may enhance the utility of prediction, and as a consequence, adaptation, is attention and reward (Gambi, this volume; Hopp, this volume). An idea that has recently become more popular is that learners may need to experience “desirable difficulties” (Bjork & Kroll, 2015; Hertel, 2020). In short, learners need to be challenged and experience errors to learn. However, the challenges need to be such that they can potentially be overcome (hence “desirable”). Potts et al. (2019) show that learners who were asked to predict a translation (and hence, could make errors) before seeing the correct translation had better retention afterwards (see also Grüter et al., this volume). As Gambi (this volume) points out, learners who are asked to explicitly predict and sometimes make errors, may also pay more attention, and may be more motivated than learners in a more passive paradigm. For learners who are explicitly predicting, adaptation may therefore lead to a higher utility of the prediction, namely the reward of making a correct prediction.

Although controversial, there is some evidence that prediction and prediction errors can contribute to L2 learning, but that factors such as attention and awareness may play an important role. More research is needed to better understand what may induce L2 learners to predict and to learn from prediction errors, and how this could ultimately be implemented in the L2 classroom.
Concluding remarks, synopses of chapters, and future directions

Research on predictive language processing in the past decade has identified what kind of cues can be used to predict various kinds of upcoming information. The critical question is now how listeners and readers achieve maximal processing efficiency: Under what circumstances do listeners and readers use which cues to predict and adapt? And when is it more efficient not to predict? This approach shifts away from the debate over how essential or marginal prediction is for language processing and learning overall. Instead, it encourages research examining under what circumstances prediction is beneficial, and how these circumstances may vary between different language users and across development.

The chapters in this volume all contribute to this endeavor in various ways. Ito and Pickering offer a theoretical framing for the study of prediction in L2 processing within production-based models of prediction (Pickering & Garrod, 2013) and outline how the largely non-automatic processes that are hypothesized to underlie prediction-by-production may account for reduced engagement in prediction in an L2, especially in domains that align with later stages in the production process (syntactic and phonological encoding, vs. semantics).
Schlenter and Felser present an empirical contribution in which they examine L2 learners' ability to use information from different linguistic domains (lexical semantics, morphosyntax) for prediction during sentence processing in two eye-tracking studies with Russian learners of German. Their results show predictive use of both verbal semantics and case marking cues; yet subtle differences between L1 and L2 speakers emerged, with the latter showing greater uncertainty and susceptibility to competition, which manifested in delayed information integration.

Chun, Chen, Liu and Chan report novel empirical data from a visual-world eye-tracking study that extends the inquiry of semantic prediction to syntactically complex sentences containing relative clause attachment ambiguities. Their findings indicate that Chinese learners of English used semantic information from the verb to predict an upcoming object in both simple and complex syntactic contexts, yet prediction effects emerged somewhat later in the complex sentence condition, suggesting modulation of semantic prediction by the cognitive load associated with the concurrent processing of complex syntax.

Foucart's chapter offers a comprehensive review and discussion of the role that cross-linguistic influence (CLI) may play in L2 predictive processing, and outlines how CLI may be integrated into current models of prediction such as error-based implicit learning and prediction-by-production. Karaca, Brouwer, Unsworth and Huettig also address the
role of CLI, yet focusing specifically on a population of language users that has received little attention in the prediction literature so far, bilingual children. Compared to monolingual children, bilingual children vary more extensively in the quantity and quality of the input they experience in each language, variability that is known to relate to the development of language skills in general, and predictive processing in particular, among monolingual children and adults. Bilingual children thus offer a unique opportunity for exploring the complex relationships between prediction, language proficiency, and language experience.

Tomić and Valdés Kroff address the role that code-switching may play as a predictive cue in bilingual processing. They summarize and discuss two studies in which Spanish-English bilinguals were presented with code-switched or unilingual sentences containing target words that differed in frequency or emotional valence. Findings indicate that experienced code-switchers can use a code switch event as a signal for upcoming information that is unexpected or unusual. These studies present a novel approach to the study of code-switching that shifts focus from the view of code-switching as a cause of integration cost to the benefits that a code-switch event might confer on processing downstream by way of prediction.

The last three chapters all focus on the relation between prediction and L2 learning. Hopp discusses the complex and reciprocal relation
between predictive processing and learning and points to potential limitations in the role that prediction may play in the L2 learning of grammatical properties. As Hopp argues, learning from prediction error can only occur if the learner is able to revise a parse after the encounter of a prediction error. If either current knowledge or resources are insufficient to arrive at an appropriate revision, learning cannot take place, and unrevised prediction errors may result in persistent misparsing and fossilization. Hopp calls for future work assessing L2 learners’ ability to benefit from prediction (error) across development and across tasks, and the role of awareness and explicit memory therein.

Drawing on insights from recent work on L1 vocabulary development in children and from research on novel word learning among adults, Gambi discusses the role that prediction may play in L2 vocabulary learning. Behavioral and neuroscientific evidence shows that guessing the meaning of an unknown word benefits the encoding of lexical knowledge in long-term memory. While these findings are consistent with error-driven learning accounts, Gambi argues that they are also compatible with explanations that point to the generation of a guess as a trigger for curiosity and motivation generated by the potential reward value of linguistic knowledge, and calls for future studies addressing the contributions of these factors in more naturalistic L2 word learning contexts.
In the last chapter, Grüter, Zhu and Jackson present findings from a structural priming experiment that examined whether forcing L2 learners to engage in prediction would lead to increased adaptation and learning. Their findings indicate that Korean L2 learners of English who had to guess how a virtual partner would describe pictures were more likely to use a construction used by that partner in their own subsequent picture descriptions than learners who only had to repeat the partner's description. Consistent with Gambi's discussion of guessing benefits in word learning, Grüter and colleagues observe that these results are consistent with error-driven learning accounts, but call for future work to examine the role of factors such as attention, awareness and motivation that may have contributed to the learning benefit observed when participants were forced to guess.

Collectively, the chapters in this volume present an impressive array of empirical and conceptual insights into the role of prediction in L2 processing and learning. Yet as they also make clear, we still have a long way to go to more fully understand the mutual relation between prediction and learning in L2. To push such research forward, we encourage longitudinal studies to investigate how prediction and its relation with learning develop over the course of learning, over the lifespan, and as a function of bilingual experience beyond global proficiency. Computational models are needed to better specify how multiple factors interact and
operate, and when the cost of prediction outweighs its benefits (e.g. Aurnhammer & Frank, 2019; Kleinschmidt et al., 2015; Martin, 2016).

Finally, research should be extended beyond traditional laboratory contexts, so as to test how current insights scale up to naturalistic conversations and real-life SLA contexts (Heyselaar et al., 2020; Schremm et al., 2017).

The goal of research on predictive processing is to account for differences within individuals, across individuals, as well as across populations. Specifying which cues are used to predict what, and determining when predicting is more efficient than not predicting, will be a big step towards that goal.

Acknowledgments

The authors would like to thank Holger Hopp for valuable feedback on an earlier draft. EK is in part supported by NSF BSC-2017251.

References


https://doi.org/10.1016/j.jml.2019.02.006


https://doi.org/10.1037/0033-295X.108.3.624


https://doi.org/10.1016/j.neuropsychologia.2019.107225


https://doi.org/10.1016/j.jecp.2017.01.007

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency,


Language Acquisition, 35, 353-387.
https://doi.org/10.1017/S0272263112000915

https://doi.org/10.1177/0963721418794491

https://doi.org/10.1016/j.jml.2013.06.001


https://doi.org/10.1111/cogs.12022


https://doi.org/10.1017/S0272263116000437

https://doi.org/10.1177/0956797615625223

https://doi.org/10.1002/dys.1497

https://doi.org/10.1016/j.brainres.2018.11.013


Second Language Acquisition, 1-22.
https://doi.org/10.1017/S0272263120000443


https://doi.org/10.1080/23273798.2015.1102299

https://doi.org/10.1162/jocn_a_00328

https://doi.org/10.3389/fpsyg.2015.00731


